

EEG-based classification of video quality perception using steady state visual evoked potentials (SSVEPs)

Laura Acqualagna^{1,8}, Sebastian Bosse^{2,8}, Anne K Porbadnigk³,
Gabriel Curio⁴, Klaus-Robert Müller^{3,5}, Thomas Wiegand^{2,6} and
Benjamin Blankertz^{1,7}

¹ Neurotechnology Group, Technische Universität Berlin, Berlin

² Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany

³ Machine Learning Group, Technische Universität Berlin, Berlin

⁴ Neurophysics Group, Charité, Berlin, Germany

⁵ Department of Brain and Cognitive Engineering, Korea University, Seoul

⁶ Department of Electrical Engineering, Technische Universität Berlin, Germany

⁷ Bernstein Focus Neurotechnology, Berlin

E-mail: laura.acqualagna@tu-berlin.de, sebastian.bosse@hhi.fraunhofer.de, gabriel.curio@charite.de,
klaus-robert.mueller@tu-berlin.de, thomas.wiegand@hhi.fraunhofer.de and benjamin.blankertz@tu-berlin.de

Received 13 October 2014, revised 22 January 2015

Accepted for publication 30 January 2015

Published 13 March 2015



CrossMark

Abstract

Objective. Recent studies exploit the neural signal recorded via electroencephalography (EEG) to get a more objective measurement of perceived video quality. Most of these studies capitalize on the event-related potential component P3. We follow an alternative approach to the measurement problem investigating steady state visual evoked potentials (SSVEPs) as EEG correlates of quality changes. Unlike the P3, SSVEPs are directly linked to the sensory processing of the stimuli and do not require long experimental sessions to get a sufficient signal-to-noise ratio. Furthermore, we investigate the correlation of the EEG-based measures with the outcome of the standard behavioral assessment. **Approach.** As stimulus material, we used six gray-level natural images in six levels of degradation that were created by coding the images with the HM10.0 test model of the high efficiency video coding (H.265/MPEG-HEVC) using six different compression rates. The degraded images were presented in rapid alternation with the original images. In this setting, the presence of SSVEPs is a neural marker that objectively indicates the neural processing of the quality changes that are induced by the video coding. We tested two different machine learning methods to classify such potentials based on the modulation of the brain rhythm and on time-locked components, respectively. **Main results.** Results show high accuracies in classification of the neural signal over the threshold of the perception of the quality changes. Accuracies significantly correlate with the mean opinion scores given by the participants in the standardized degradation category rating quality assessment of the same group of images. **Significance.** The results show that neural assessment of video quality based on

⁸ Both authors contributed equally to this work.



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

SSVEPs is a viable complement of the behavioral one and a significantly fast alternative to methods based on the P3 component.

Keywords: EEG, SSVEPs, video quality assessment, classification, MOS

(Some figures may appear in colour only in the online journal)

1. Introduction

Perceived quality of a set of images or videos is usually assessed using some opinion tests, in which participants are asked to rate the quality of a given visual stimulus on a rating scale (ITU 2002, 2008). This type of behavioral tests has some well known limitations: first, they require a large number of participants to obtain statistical significance and the results often suffer from large inter-subject variance. Second, the judgment can be biased by several factors not depending on the quality of the stimuli, such as the nature of the task, the rating scale used, the internal state of the participant, expectations, emotions. A possible way to quantify visual distortions more objectively would be modeling some key features of the visual system (Jayant *et al* 1993, Seshadrinathan and Bovik 2010). Visual perception is a complex process. The visual stimulus passes through the optical system and excites the photoreceptors of the retina. The resulting signal is transferred to the visual cortex and processed following further elaboration by higher levels of the visual system. Humans are assumed to have an internal threshold, which makes them decide at the cognitive level whether they noticed a distortion, or not. So far, a precise model of subjective perception is not yet available and the process underlying the judgment itself is not fully understood. Therefore, in recent years, there has been an increasing interest to assess perceived image quality through the direct measurement of brain signals. For this purpose, one technology that proved to be suitable is electroencephalography (EEG), which records the ongoing electrical brain activity. EEG has been widely exploited in visual and audio perception research (Babiloni *et al* 2006, Demiralp *et al* 2007, Busch *et al* 2009, Porbadnigk *et al* 2013) and recently also for assessing user's perceived multimedia quality (Hayashi *et al* 2000, Porbadnigk *et al* 2010, 2011, Lindemann *et al* 2011, Lindemann and Magnor 2011, Antons *et al* 2012, Mustafa *et al* 2012, Scholler *et al* 2012, Perez and Delechelle 2013, Porbadnigk *et al* 2013, Moldovan *et al* 2013, Arndt *et al* 2014, Bosse *et al* 2014, Kroupi *et al* 2014). Most of these studies focus on the measurement of specific EEG components called event related potentials (ERPs). ERPs are brain responses arising time locked to the onset of an external significant stimulus (Donchin 1979, Pfurtscheller and Lopes da Silva 1999, Picton *et al* 2000). There are several categories of ERPs, associated to different steps of the processing of the stimulus, with different latencies and scalp topographies. In video quality assessment studies, the most exploited ERP component has been the P3, which is a positive deflection of the brain activity arising between 300 and 500 ms after the

stimulus onset in the central-parietal cortex (Smith *et al* 1990, Picton 1992, Johnson 1993), largely independent on the sensory modality. In Scholler *et al* (2012), a P3-based EEG measurement is used to directly assess the perception of quality changes in 8 s video clips, in which distortions are introduced by a hybrid video codec. They find that quality changes elicit a P3 component that is positively correlated with the magnitude of the change, which can be classified on a single-trial basis. Also they report that a participant shows brain responses to low distortion stimuli, although the participant did not report perceiving them. In another study by Lindemann and Magnor (2011), the use of EEG as a tool for evaluating image quality for JPEG-compressed images is investigated. They show that the presence of JPEG artifacts elicits ERPs whose amplitudes vary with the compression ratio. In a follow-up study (Lindemann *et al* 2011), they present a first evaluation of artifacts in simple video stimuli to verify whether they can be assessed using ERP analysis. By comparing the ERPs elicited by videos with and without two types of artifacts, they show that artifacts produce measurable ERPs whose shape depends on the strength of the artifacts. Another approach for single-trial classification of artifacts in videos is proposed by Mustafa *et al* (2012), where the focus is the classification of artifacts that usually occur in image-based rendering techniques. They also show that it is possible to distinguish with a certain degree of accuracy between different types of artifacts. All these works demonstrate that an EEG-based approach in classifying quality changes or artifacts in videos and images is a feasible measurement that, together with behavioral tests, can lead to a more objective rating of perceived quality. Notwithstanding, all studies that capitalize on the P3 component have several limitations. First, the P3 component is not directly linked to sensory processing, but reflects higher cognitive processing of the stimulus. It is elicited by an oddball paradigm, in which the user pays attention to the occurrences of a target event among more frequent non-targets. The paradigm usually requires a high number of trials in order to have a sufficient number of target events and a good signal-to-noise ratio. Besides, the design of such experiments needs to be carefully done. For example, the target events have to be sufficiently distant in time between each other, in order to produce the 'surprise effect' which would significantly modulate the brain signal (Duncan-Johnson and Donchin 1977, Sellers *et al* 2006).

In this study, we investigated an EEG-based approach to video quality assessment, exploiting steady state visual evoked potentials (SSVEPs). SSVEPs are oscillatory brain responses elicited in the visual cortex by a repetitive visual stimulus (Calhoun and McMillan 1996, Herrmann 2001,

Vialatte *et al* 2010). For example, a flickering light would synchronize the neurons of the visual cortex at the same frequency of the driving stimulus and higher harmonics (Regan n.d., year, Rager and Singer 1998). They have been employed successfully in brain computer interfaces (BCIs) (Gao *et al* 2003, Müller-Putz *et al* 2005, Friman *et al* 2007, Allison *et al* 2008, Parini *et al* 2009) and in basic research on the human visual system (Morgan *et al* 1996, Müller and Hübner 2002, Keil *et al* 2003, Pastor *et al* 2003), proving to be robust and reliable signals to classify. Previous work on using SSVEPs has already demonstrated the basic suitability of the targeted method for video quality measurement (Norcia *et al* 2014). Using a SSVEP-based paradigm in which the flickering effect is caused by changes in video quality (owing to compression), we show that it is possible to elicit a direct visual response and to obtain a meaningful quantification by single-trial classification using machine learning techniques. A characteristic that makes SSVEPs preferable to P3 in video quality assessment is that SSVEPs are directly linked to the sensory processing and can give a more straightforward indication of the level of perception of the visual stimuli. Moreover, for the reasons described above, a P3-based approach requires long inter (target-) stimulus intervals in order to achieve a suitable signal-to-noise ratio, while a SSVEP-based approach allows to collect a large number of trials within a short amount of time. For example, in the study of Scholler *et al* (2012), a total of 600 ERPs was collected in 120 min, while more than 10,000 ERPs (single VEPs of the SSVEPs) can be collected using the proposed SSVEP-based approach within the same amount of time. Of course, such a comparison can only give an indication of the potential to speed up the assessment by using SSVEPs, since comparing just the number of trials ignores that both studies had different stimulus material and presumably a different signal-to-noise ratio in single-trials. Another limitation of previous studies is that they do not investigate directly the correlation between the neural assessment and the behavioral ratings. An exception can be found in the recent study of Arndt *et al* (2014), in which they performed a series of experiments assessing video and audiovisual quality degradation. They show an average significant correlation between the P3 amplitudes and the mean opinion scores (MOS) in three out of five experiments. In this study, we performed the standard behavioral test for video quality assessment either before or after the EEG measurement and the MOS for the same groups of textures was collected. Therefore, we could correlate the classified SSVEP features not only with the quality changes intrinsic to the video signals, but also with the judgments of the participants. We used textures in condition of natural luminance as stimuli instead of artificially generated stimuli. This makes the stimuli more realistic and the experimental condition closer to real world video sequences. Yet, the textures were simple enough to avoid influences of semantic content, and chosen to be homogeneous in order to minimize the effects of eye movements and to avoid that attention was captured by salient objects not related to the purpose of the experiment.

2. Methods

2.1. Stimuli

Six gray-level texture images (Ojala *et al* 2002, Kyllberg 2011) were chosen as the basis for stimulus generation (figure 1). The size was 512×512 pixels and they all had the same mean luminance. In order to make the measurement independent of the image statistics and of the actual gaze position during the experiment, the texture images were spatially roughly stationary. The quality of each texture image was then degraded in six different levels. The distortions were introduced by coding the textures using the HM10.0 test model (JCT-VC 2014) of the emerging high efficiency video coding (H.265/MPEG-HEVC) standard (Sullivan *et al* 2012). In this standard, statistical redundancies are exploited by block-wise temporal and spatial linear prediction. The residual signal is transformed block-wise, and coefficients are quantized in the transform domain. The quantization is controlled by the quantization parameter (QP). Coding artifacts, which are perceived by the human observer as a loss of visual quality, are introduced by the quantization of the transform coefficients. In order to investigate how the visual cortex responds to distortions at the threshold of perception, the first three distortion levels are chosen to be perceived as high quality. The QP-values used in the experiment were estimated in a pilot study in order to meet consistent MOS. All the texture images in all the different levels of degradation were displayed as videos, 114 s long. Details about the structure of the videos are described in section 2.2.3.

2.2. Experimental design

2.2.1. Participants. Sixteen participants (seven females and nine males, in the age group 21–46) took part in the experiment. All had normal or corrected-to-normal vision and none of them had a history of neurological diseases. They were all native German speakers or at least with a level of German comprehension of five, on the six level scale of competence laid down by the Common European Framework of reference for languages (Little 2007). All of them were naïve in respect of video quality assessment studies and were paid for their participation. The study was performed in accordance with the declaration of Helsinki and all participants gave written informed consent.

2.2.2. Apparatus. EEG was recorded with sampling frequency of 1000 Hz using BrainAmp amplifiers and an ActiCap active electrode system with 64 channels (both by Brain Products, Munich, Germany). The electrodes used were Fp1,2, AF3,4,7,8, Fz, F1-10, FCz, FC1-6, FT7,8 Cz, C1-6, T7, CPz, CP1-6, TP7,8, Pz, P1-10, POz, PO3,4,7,8, Oz, O1,2. The electrode that in the standard EEG montage is placed at T8 was placed under the right eye and used to measure eye movements. All the electrodes were referenced to the left mastoid, using a forehead ground. For offline analyses, electrodes were re-referenced to linked mastoids. All impedances were kept below 10 kOhm. The stimuli were

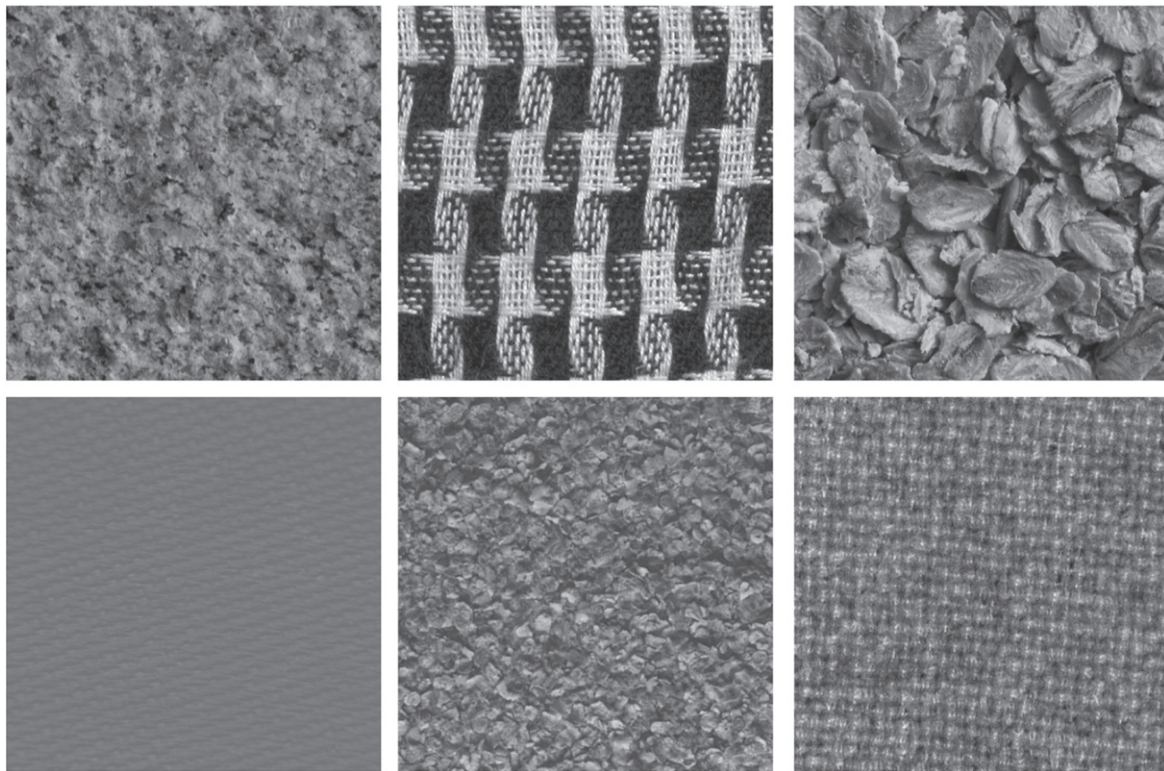


Figure 1. Stimuli. Six natural images were chosen as a basis for stimuli generation. The textures in the upper row represent stone, scarf and oatmeal, in the lower row gray rubber, gray flakes and blanket, all in their undistorted form. They have been degraded in six levels of quality coded with the HM10.0 test model HEVC standard and grouped together to form videos with a frame rate of 3 Hz.

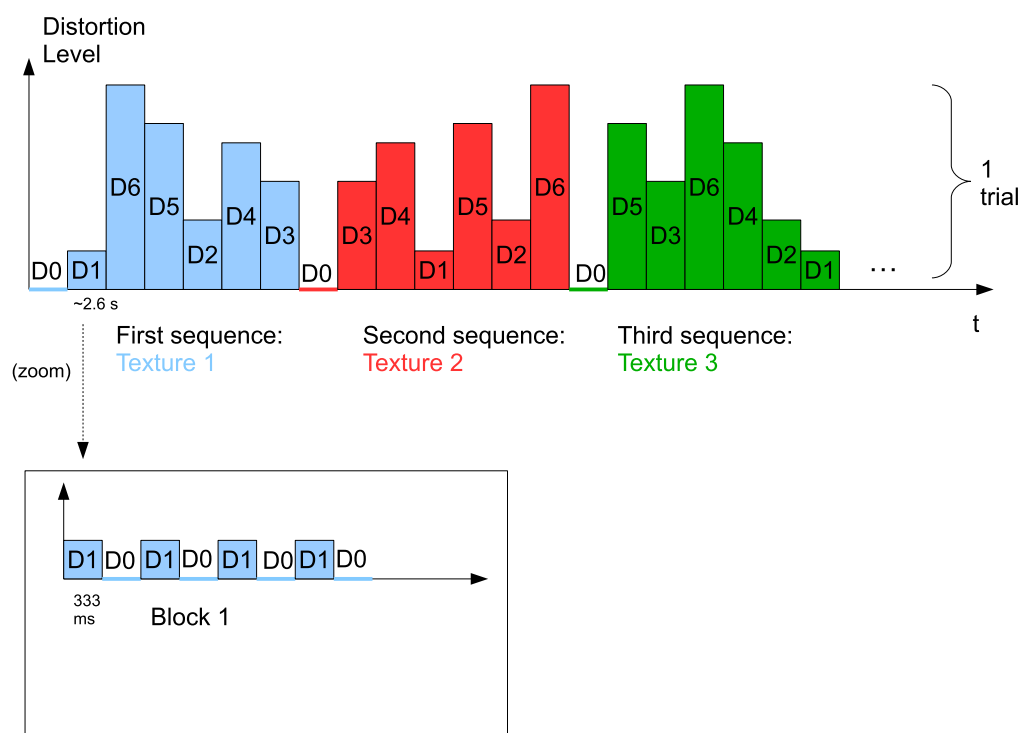


Figure 2. Video structure. Each video (trial) comprised the six textures presented in all the levels of distortion (D1, ..., D6) in a random order. Each texture was displayed distorted for 333 ms, followed by the undistorted form for 333 ms (D0) and the same succession was repeated four times for each level. Such alternation of quality changes produced a flickering effect eliciting SSVEPs when perceived by the participant.

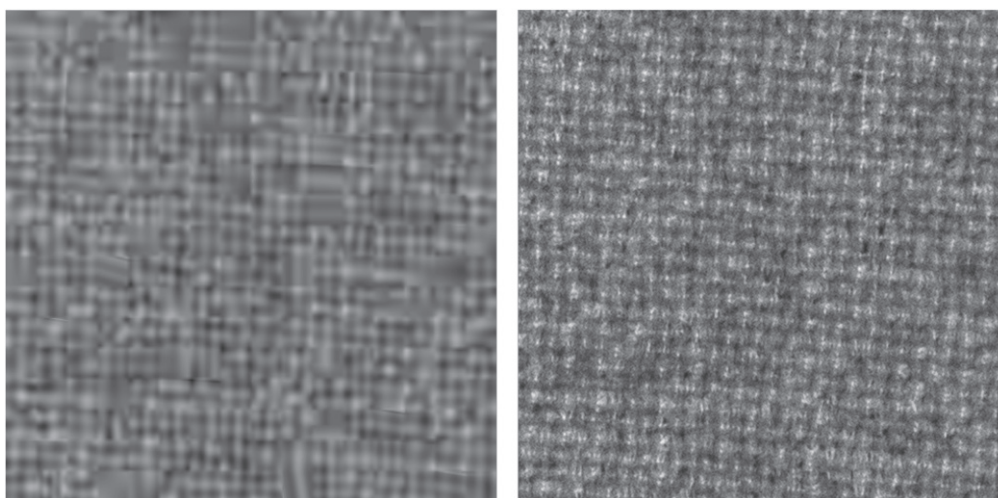


Figure 3. Behavioral assessment. Textures were presented in a random order, the distorted form on the left side of the display and the undistorted on the right (here an example of blanket with degradation level D6). Participants had 10 s to look at them in free viewing condition before the evaluation of the quality level. Each level of distortion was presented for three times, for each texture.

shown on a 23" screen (Dell U2311H) with a native resolution of 1920×1080 pixels at a refresh rate of 60 Hz. The screen was normalized according to the specifications in ITU (2002). The stimuli resolution was 512×512 pixels (128×128 mm), which corresponds to 7.15 visual angle. The size of the images in the behavioral part of the experiment was the same as in the videos. The viewing distance was 110 cm, in compliance with specifications in the ITU-T Recommendation P.910 (ITU 2008). Subjects sat in front of the display in a dimly light room.

2.2.3. Procedure. The experiment consisted of an EEG measurement and a behavioral part. After a general introduction to the experiment and the preparation of the EEG cap, half of the participants started with the EEG recording and half with the behavioral assessment. In the EEG part, they had to watch a series of 51 videos, divided into three runs ($20 + 15 + 16$). Between each run, there was a break of about 10 min, to give the participants the chance to relax and stretch. Each video had a length of 114 s, followed by a pause of 5 s. Each video began with a fixation cross that was displayed for 3 s at the center. In order to minimize artifacts, participants were instructed to not move their eyes during the presentation of the video and to blink as little as possible. Figure 2 depicts the structure of one video, which will be addressed as 'trial' in the remainder of the paper. Each trial comprises all six textures and degrees of quality, presented in random order. The stimulus onset asynchrony (SOA) was 333 ms, that is, each texture image was displayed for 333 ms. At the beginning, the texture was presented in its undistorted form (D0) for 2664 ms ($333 \text{ ms} \times 8$). Then, the first quality change occurred and the distorted texture was displayed for 333 ms, followed by 333 ms of the same texture in its undistorted version. This cycle 'distorted-undistorted' was repeated four times for a total of 2664 ms. Then, the same texture was displayed with another level of quality change,

for a cycle 'distorted-undistorted' of the same length. This procedure was performed until all the distortion levels were displayed for that texture (randomized order). After that, the texture was switched and the new one was displayed at the beginning in its undistorted version for 2664 ms, before starting the cycle of quality changes. This presentation elicits SSVEPs if the changes due to altered quality are processed in the visual cortex. A SOA of 333 ms results in a flickering frequency of 3 Hz. Before starting the main EEG recording we performed some additional measurements, comprising a relax measurement and an artifact measurement. In the relax measurement, EEG was acquired during rest alternating with eyes open and eyes closed (10 s each) in order to obtain a standard measure of the participant's occipital alpha rhythm. In the former phase, they had to look at a simple colored geometrical shape moving in the center of the display. The cycle 'eyes closed-eyes open' was repeated ten times. In the artifact measurement, five crosses were displayed: one in the center, the others respectively at the left and right side, upper and below the central one. The distance of the four external crosses matched the size of the videos. Participants were instructed to fixate the central cross and then promptly move the eyes to one of other four, according to the instructions of a recorded voice. In the behavioral part of the experiment, participants had to evaluate the perceived quality of the textures, following the standardized degradation category rating quality assessment (ITU 2002) in a presentation mode. Each texture was presented in the display in pairs for 10 s: on the right-hand side in its original undistorted version (reference) and simultaneously on the left-hand side with changed quality (figure 3). After that, a new window was displayed, with the nine-grade degradation (distortion) scale, according to the ITU-T P-910 recommendation (ITU 2008). The scale was displayed in German language, and in English can be translated in the following: 1- very annoying; 3- annoying; 5- slightly annoying; 7- perceptible, but not annoying; 9- imperceptible. Grade 8 is commonly

interpreted as the perceptibility threshold, that is the distortion level where the observer is not completely sure to perceive the distortion. Participants had up to 10 s to decide about the level of distortion of the previously displayed left image compared to the reference one on the right, scrolling a bar until the selected grade and confirming by button press. After that, the presentation switched to the next pair of textures. If the person did not make any choice within the 10 s, the presentation automatically went ahead with the next comparison. In the behavioral assessment, each texture image was presented in all the level of distortions (comprising the comparison reference–reference). For each level, there were three evaluations. The order of the evaluations was randomly shuffled. At the beginning of the assessment, a calibration block was displayed in order to make the participants confident with the test: each texture was displayed for just two evaluations, worst quality level versus reference and reference versus reference, for a total of 12 calibration evaluations. Like in the actual behavioral assessment, in this short test participants were not aware of the quality level of the displayed textures. The data of this calibration block were not considered in the analysis.

2.3. Preprocessing and data analysis

EEG signal was lowpass filtered from 0 to 40 Hz with a Chebyshev filter of order ten (3 dB of ripple in the passband and 40 dB of attenuation in the stopband) and down-sampled to 100 Hz. For the SSVEP visualization, the continuous signal was divided into epochs ranging from 0 to 2664 ms, relative to the onset of the first distorted texture for each quality change. Each epoch comprises all four repetitions of the same quality level. Epochs referring to the same distortion level were averaged over all the trials and all the textures. In the frequency-domain analysis, EEG spectra were calculated between 1 Hz and 18 Hz for all epochs and then averaged for each distortion level, over all trials and textures. For single-trial offline classification, we tested two different methods. The first one exploited the oscillatory nature of the SSVEPs and used common spatial pattern (CSP) analysis to enhance the discrimination of the event-related synchronization (Koles *et al* 1990, Ramoser *et al* 2000, Blankertz *et al* 2008, Parini *et al* 2009). The second one is based on methods used in ERP analysis and exploits spatio-temporal features (Blankertz *et al* 2011).

2.3.1. CSP method. CSP was used to extract spatial filters in order to enhance the signal of interest in the occipital cortex. In general, CSP aims at maximizing the variance on the spatially filtered signals under one condition while minimizing it for the other condition. In our case, one condition would be the neural signal associated to the distorted images and the other condition would be that corresponding to the undistorted images presented at the beginning of each block of textures. Since variance of bandpass filtered signals is equal to band-power, CSP analysis is applied to band-pass filtered signals in order to obtain an effective discrimination of mental states between the two

conditions. CSP projects the signal $\mathbf{x}(t) \in \mathbb{R}^C$ in the original sensor space to $\mathbf{x}_{\text{CSP}} \in \mathbb{R}^C$, which lives in the surrogate sensor space, as follows:

$$\mathbf{x}_{\text{CSP}}(t) = \mathbf{W}^T \mathbf{x}(t). \quad (1)$$

Each column vector $\mathbf{w}_j \in \mathbb{R}^C$ ($j = 1, \dots, C$) of \mathbf{W} is a spatial filter; each column vector $\mathbf{a}_j \in \mathbb{R}^C$ ($j = 1, \dots, C$) of a matrix $\mathbf{A} = (\mathbf{W}^{-1})^T \in \mathbb{R}^{C \times C}$ is a spatial pattern. While for classification only the spatial filters are used, only the patterns allow for a physiological interpretation of the CSP components, see Blankertz *et al* (2008) and Haufe *et al* (2014). For a more detailed review on CSP analysis and its application to EEG signal processing, please refer to Lemm *et al* (2005), Lotte *et al* (2007), Blankertz *et al* (2008), Sannelli *et al* (2011), Samek *et al* (2012). In our case, the CSP filters were calculated between the epochs of maximum distortion level, which will be named class D6, and epochs of the undistorted level, which will be named class D0. Since the performance of this spatial filter depends on the operational frequency band of interest, manually selecting a specific frequency range is commonly used with the CSP algorithm (Dornhege *et al* 2006, Ang *et al* 2008). In our case, CSP was performed after filtering the data with a 5th order Butterworth filter centered at 3 Hz (pass-band 2–4 Hz), 6 Hz (pass-band 5–7 Hz) and both simultaneously (filter bank). The filter bank concatenated data filtered at 3 and 6 Hz, which were subsequently spatially filtered with the respective CSPs. In this way, features referring to both frequencies could be exploited simultaneously. In all the three cases, the continuous EEG signal was divided into epochs of 667 ms length, time-locked to the onset of the distorted image. For each texture and level of distortion, the four repetitions of the cycle ‘distorted-undistorted’ were averaged. For D0, the first block of epochs at the beginning of each video was discarded, because it is often affected by artifacts due to subject’s movements between the videos. For training and testing, the epochs were split into a subset of epochs with even IDs (training) and odd IDs (testing), in order to prevent that longer-term changes in the EEG during the experiment could bias classification. For the calculation of the CSP filters (training), the epochs with even IDs were considered. One up to three CSP filters per class were automatically selected for each subject and checked visually. The CSP filters that maximize the variance for class D6 while minimizing the variance for D0 were used. The selected filters were then applied to the training epochs comprising D6 and D0 and the log-variance in three equally spaced intervals of the filtered data was used as feature matrix for training a classifier based on linear discriminant analysis (LDA) (Lemm *et al* 2011). The CSP filters were then applied to the testing epochs, from D1 to D6, and classification was made between each level of distortion and D0. Classification performance was measured by the area under the curve (AUC) of the receiver operating characteristic (Hanley and McNeil 1982).

2.3.2. Spatio-temporal features method. In this method, we classified single-trial visual evoked potentials time-locked to

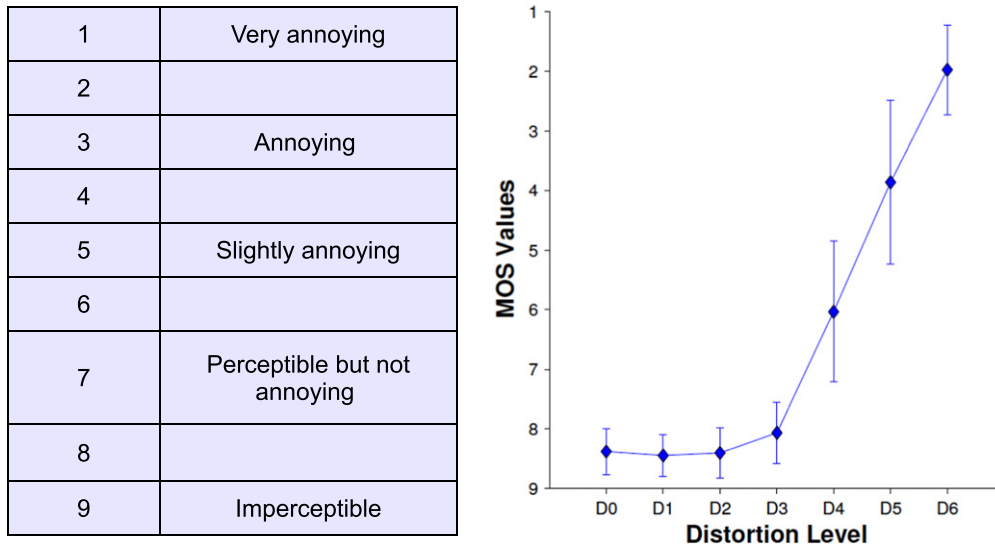


Figure 4. Behavioral assessment. Left: table representing the nine level scale of scores that the participants could give to the quality of the distorted texture compared to the undistorted. Right: mean opinion scores given by participants for each quality level. Error bars refer to the standard deviation from the mean. Until level D3, mean MOS values remain above or at the level of perception threshold (level 8) and from D4 they decrease linearly with increasing distortion.

the onset of the distorted texture, using spatio-temporal features (Tomioaka and Müller 2010, Blankertz *et al* 2011). The EEG data was divided into epochs as follows: for the ‘class 1’ events, epochs were time-locked to the onset of the distorted textures. For the ‘class 2’ events, the starting point of epochs was shifted 160 ms after the onset. Being the period of the oscillatory visual response of about 333 ms, a shift of 160 ms would lead to a high discrimination between the two classes. In both classes, the length of the epochs was 667 ms. For feature extraction, five temporal windows were selected individually for each participant by a heuristic (Blankertz *et al* 2011), based on the pointwise biserial correlation coefficient (r -values). More specifically, we used the sign $-r^2$ as a measure of separability between the two classes. The aim of the heuristic is to find time intervals that have a fairly constant pattern (class 1 minus class 2 difference) and maximal r^2 differences. Features were calculated from 36 channels (FC1,3,z,2,4, C1,3,5,z,2,4,6, CP1,3,5,z,2,4,6, P1,3,2,4 Pz,7,9,8,10, PO3,7,4,8, O1,z,2) by averaging voltages within each of the five chosen time windows resulting in $36 \times 5 = 180$ dimensional feature vectors. Classification was performed using ten-fold cross-validation and LDA with shrinkage of the covariance matrix (Blankertz *et al* 2011).

Classification results were correlated with the average magnitude of the alpha rhythm during the experiment. The alpha peak was searched within a range of frequencies. In order to estimate alpha power, the difference between the value of the alpha power and the linear interpolation between the flanking frequencies (as baseline) was calculated. For this estimate, flanking frequencies have been searched within the intervals 6–10 Hz and 11–13 Hz, and the alpha peak was determined as maximum of the spectra between the flanking frequencies. For all participants, alpha peak was detected between 9 and 12 Hz.

3. Results

3.1. Behavioral data

Figure 4 displays on the left the nine-degradation scale of the MOS, in which grade 8 corresponds to the perception threshold of the degradation, that is the level where the observer is not completely sure anymore to perceive any degradation. On the right, the mean MOS values over all the participants, textures and repetitions is plotted as a function of the distortion level. The error bars represent the standard deviation as a measure of the variation of the mean MOS across participants. The plot displays that on the behavioral level the participants were not able to discriminate the two most subtle distortion levels (D1 and D2) from the reference image (D0). For level D3, the mean MOS is lower but still remains above the value of the perception threshold. From distortion level D4 to D6, the mean MOS values decrease linearly with the level of distortion, and the error bars also display an increase in variability among participants.

3.2. Neurophysiological data

Figure 5 (left) displays representative SSVEP waveforms of participant VPib recorded at an occipital scalp site (electrode position Oz). The plots are ordered according to increasing distortion levels, that is, the top row refers to the reference texture of highest quality (D0) and the bottom row to the maximum distortion level (D6). Each plot represents the average EEG activity over all the trials and textures at that specific quality level. The time zero is locked to the onset of the first repetition of the texture of each block. On the x axes, Dist refers to the onset of the distorted texture, and Ref to the onset of the reference image (D0). As described in the methods section, the time interval between the onset of each

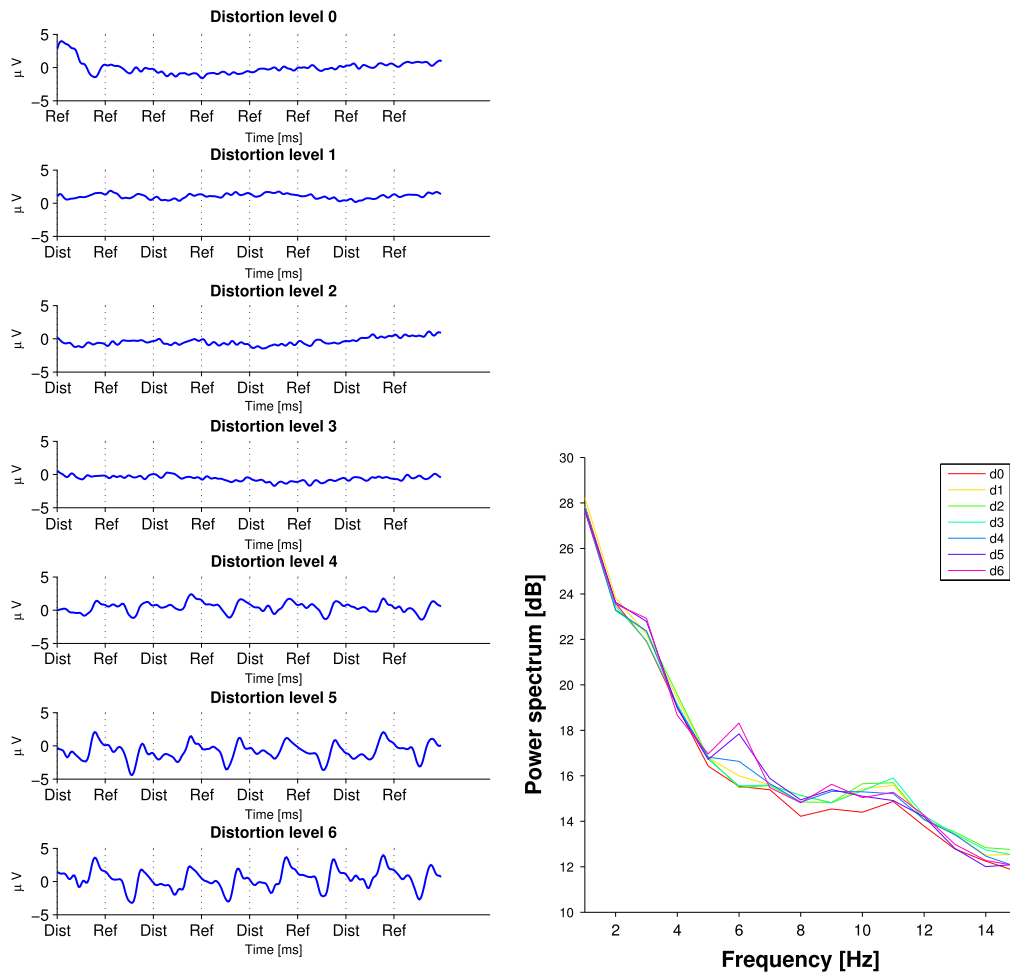


Figure 5. SSVEPs. Left: brain activity of participant VPib averaged over trials and textures at channel Oz is displayed from D0 to D6. *Dist* refers to the onset of the distorted texture, *Ref* to the undistorted. Until level D3 there is no clear modulation of the neural signal by the quality changes. From level D4 SSVEPs are visible with amplitude increasing with increasing distortion level. Right: spectra of the brain activity of participant VPib at channel Oz averaged over trials and textures. As in the time domain analysis, no clear modulation is visible until D3. From D4 to D6 (blue, violet and magenta lines), peaks at 3, 6 and 9 Hz are evident.

frame is 333 ms. For the first three levels of distortion (D1–D3), which are around the perception threshold, the ongoing EEG activity is not visibly modulated by the quality change. From D4 onwards, the SSVEPs become clearer and their amplitude increases with increasing distortion. In all the plots where the SSVEPs are evident, it can be noticed that the onset of the reference images (*Ref*) elicits a more pronounced negative peak than the onset of the distorted ones, at the same latency. This result suggests that the transition *Dist-Ref* has a stronger impact on the modulation of the visual evoked potentials than the transition *Ref-Dist*. The results of the analysis in the frequency domain analysis of the SSVEPs (electrode position Oz) of the same participant are displayed in figure 5 (right). Power spectra were calculated on single trials before averaging over the trials and textures. The plots represent the average power for each level of distortion, coded with different colors. For the first three levels of distortion (yellow, light green, emerald green) there is no clear increase of the power in any of the frequencies of interest (3 Hz and higher harmonics). From D4 to D6 (blue, violet, magenta), the spectra display two clear peaks at 3 and 6 Hz, whose

amplitudes increase significantly with increasing distortion level ($p < 0.01$). At D5 and D6 a small peak at 9 Hz becomes visible, but this modulation is much smaller compared to the first two harmonics (therefore not taken into account in further analysis).

3.3. Classification

CSP filtering is a gold standard of the processing of EEG oscillatory signals in BCIs. In the same way, spatio-temporal features have been successfully used in the analysis of ERPs (Blankertz *et al* 2011, Lemm *et al* 2011). We considered both processing methods in combination with LDA (with shrinkage when necessary) in order to exploit both natures of the SSVEPs.

3.3.1. CSP method. Figure 6 displays the results of the CSP analysis as color coded scalp topographies (for one representative participant, VPib). Figure 6 on the left refers to data filtered around 3 Hz, and in the middle to data filtered around 6 Hz. The upper plots display the spatial filter

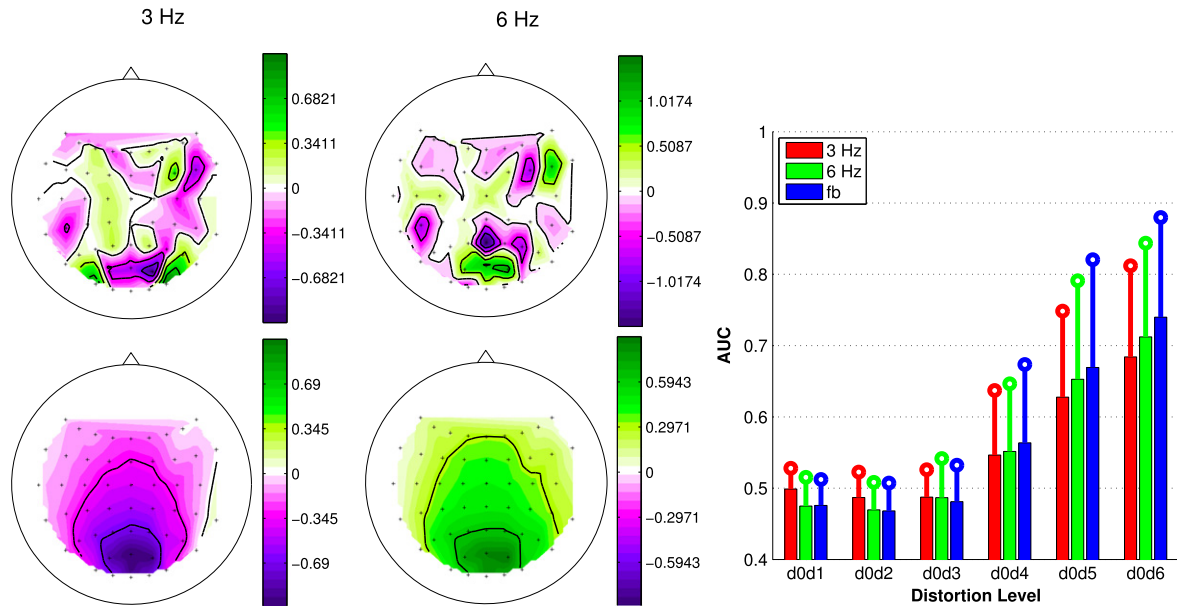


Figure 6. CSP analysis. Left and middle: scalp plots display the CSP filters (upper row) and patterns (lower row) of participant VPib, for brain activity filtered around 3 Hz (left) and 6 Hz (middle). CSP filters were calculated considering a subgroup of epochs referring to D6 and D0. The patterns clearly display that the highest variance of the neural signal takes place in the occipital cortex, where the visual information is processed. Right: average classification accuracies using LDA after CSP filtering, for features at 3 Hz (red), 6 Hz (green) and filter bank (blue), respectively. Error bars refer to the standard deviation of the participants accuracies from the mean. Classification accuracy increases significantly from D4 to D6, but no significant difference is found in the choice of the filtering frequency.

coefficients, that is, the interpolation of the values of the components of the vector w_j , the j th columns of W , at each electrode position. The bottom rows represent the corresponding pattern of activation in the brain, that is, the interpolation of the components of a_j , the j th column of $A = (W^{-1})^T$. For each participant, only those CSP filters were selected which maximize the variance of D6 while minimizing the variance of D0, since we are interested in finding the spatial patterns which discriminate the activity elicited by the distorted texture versus the reference. (Note that we use a colormap that has no direct association to signs because the signs of the vectors are irrelevant in our analysis.) After the training of the LDA classifier and the application of the CSP filters to the testing data, the classifier was evaluated for all the levels of distortions. The results are displayed in the bar plot in figure 6 (right), for 3, 6 Hz and the filter bank which considers both frequencies. The error bars indicate the standard deviation of the single participants with respect to the mean accuracy. The x axes represent the pair of classes between which the classification was performed, and the y axes the classification performances. Independent of the chosen frequency, the mean classification accuracy for the first three levels of distortion (D1 to D3) is around chance level and not affected by the quality change. Between D4 and D6, the mean classification accuracy increases linearly and significantly with increasing distortion, reaching at D6 0.68 (SD = 0.13) for 3 Hz, 0.71 (SD = 0.13) for 6 Hz and 0.74 (SD = 0.14) for the filter bank. In the CSP method a slight increase on average performances is visible when data are preprocessed with the filter bank, exploiting the features and filters of both the frequencies of interest. Repeated

measurement of ANOVA⁸ run on the AUC values with factors 'distortion level' and 'frequency' display a statistically significant difference for the first factor ($p < 0.01$), but not for 'frequency' ($F = 0.57$, $p = 0.56$). Even if the results show higher average accuracy when considering the filter bank of 3 and 6 Hz, the choice of the discriminant frequency does not affect significantly the classification accuracy.

3.3.2. Spatio-temporal features method. In the second method of classification based on spatio-temporal features, epochs of 667 ms length were considered time-locked to the onset of the distorted images for class 1, and with a lag of 160 ms for class 2. Figure 7 (left) displays the grand average of evoked potentials over all participants and trials for distortion level D6. The line colored in magenta displays the EEG activity referred to class 1, the gray line to class 2. Considering just the magenta plot, a first negative deflection is visible between 150 and 210 ms after the onset of the distorted image, followed by a pronounced positive peak around 270 ms. As already displayed in figure 5, the onset of the reference texture at 333 ms elicits a more pronounced negative visual evoked potential than that elicited by the onset of the distorted texture. The scalp plots underneath display the topographies in two time intervals where the class difference is large (shaded gray in the plot above). They visualize the distribution of the signed r^2 values as a measure of discriminability between classes 1 and 2, which is highest

⁸ Data distribution was checked with the one-sample Kolmogorov-Smirnov test under the null hypothesis that the samples are drawn from a standard normal distribution. The null hypothesis was not rejected at the 5% significance level.

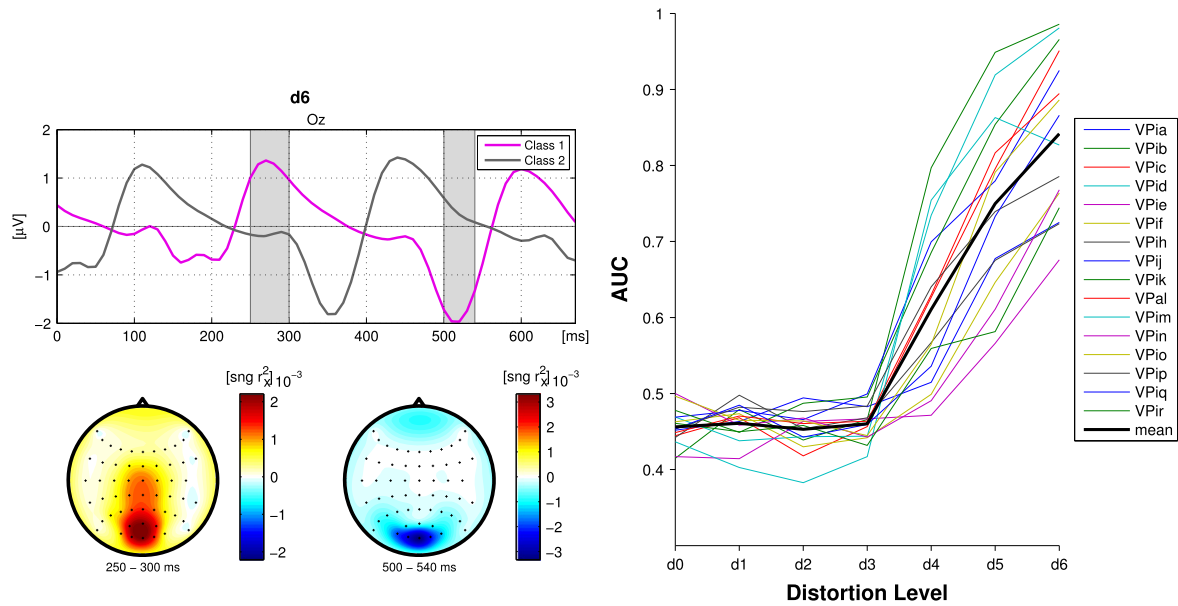


Figure 7. Classification based on spatio-temporal features. Left: grand average brain activity over all participants at channel Oz, at maximum distortion level D6. The magenta line represents class 1 and the gray line class 2. Scalp plots underneath refer to the shaded areas in the time plot and display the magnitude of the sign $-r^2$ for each channel. The highest discrimination between the two classes takes place in the occipital cortex. Right: classification accuracies using shrinkage LDA for all participants (colored lines) and mean (black thick line). Classification remains at chance level for all participants until D3, and then increases significantly with increasing distortion level. Participant VPib reaches the highest accuracy, 0.99 at D6.

in the occipital cortex with focus around the central channels. In other words, the visual processing of the stimuli leads to maximal discrimination between good and distorted quality. Based on the signed r^2 values, features for offline single-trial classification were determined (ten-fold cross-validation). A mean accuracy of 0.84 (SD = 0.1) is achieved at the maximal distortion level D6. At D5 and D4 the mean accuracy drops to 0.75 (SD = 0.12) and 0.61 (SD = 0.1) respectively. The mean accuracy at D3 down to D0 are around chance level. Figure 7 (right) displays the trend of the classification accuracy for all participants (colored lines) and the mean (black thick line) as a function of the distortion level. From distortion D3 upwards, the mean accuracy increases significantly with the level of distortion ($p < 0.01$). Since classification is based on the spatio-temporal features derived from the evoked potentials, this trend tightly follows the trend of the modulation of the visual evoked potentials in the occipital cortex. The SSVEP plots in figure 5 refer to participant VPib, who reaches the highest classification accuracy in the spatio-temporal classification method, that is 0.99, and also in the CSP method, that is 0.96 (filter bank). Participant VPib shows clear visual evoked potentials and spectra peaks at both 3 and 6 Hz.

MOS values significantly linearly correlated with classification accuracy obtained using the spatio-temporal features for all the participants ($p < 0.01$). MOS values were also correlated with the accuracy obtained by the CSP method with filter bank, which takes into account contemporary both

the frequencies of interest. In this case, we found a significant linear correlation ($p < 0.05$) for all but two participants.

3.4. Understanding individual differences

While for the first three levels of distortion there are no substantial inter-participant differences in classification results, the variability of the classification accuracy at higher levels of distortion becomes statistically significant ($p < 0.05$). For example, in the classification based on spatio-temporal features five participants reach a mean classification accuracy of more than 0.9 for the maximal level of distortion D6, while five participants never exceed 0.7. The latter even do not pass chance level in the classification of distortion level D4, thus seeming less sensitive in general to changes of visual quality. In order to find neurophysiological correlates of classification performance, we calculated the average peak magnitudes of the alpha rhythm during the experiment for each participant and correlated it with the classification result that the participant reached at the maximal level of distortion. The results are displayed in figure 8 as scatter plots, for the two classification methods described in the previous section. Each dot represents a participant. The black line is the result of the linear regression between the average magnitudes of the alpha peaks of each participant and their classification accuracies. The yellow dots represent participants having the 10% largest Mahalanobis distances to the data center, and therefore considered as outliers and removed from the analysis (Huber and Ronchetti 1975). In both classification methods used, we

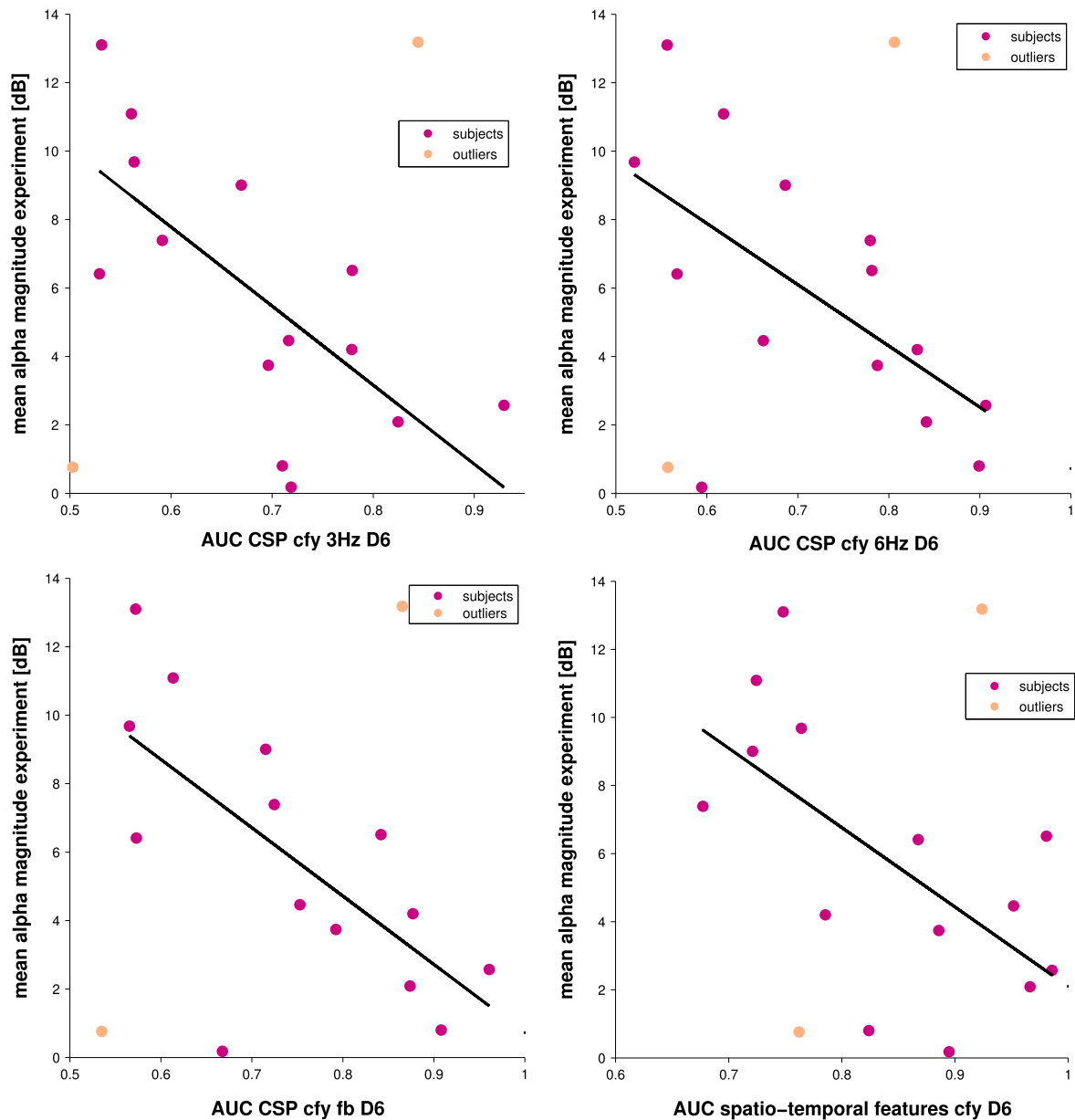


Figure 8. Correlation amplitude of alpha rhythm-classification accuracy at D6. The upper plots refer to the classification accuracy based on the CSP method at 3 Hz (left) and 6 Hz (right), the lower plots to CSP with filter bank (left) and to classification based on spatio-temporal features (right). Each dot represents one participant, the yellow ones refer to participants considered as outliers. Regression lines are also displayed. For all the methods the amplitude of the average alpha rhythm during the experiment significantly negatively correlated with the average accuracy at D6. This suggest that the alpha rhythm plays a role in the maximum classification accuracy achieved based on SSVEPs.

found a significant linear negative correlation between the average alpha activity during the experiment and the accuracy of classification of the quality change. In the CSP analysis, this is the case for all spectral features we considered (3, 6 Hz, filter bank). The Pearson correlation coefficients were $r = -0.64$ for classification based on spatio-temporal features ($p < 0.02$), $r = -0.70$ for CSP classification at 3 Hz ($p < 0.01$), $r = -0.60$ for the CSP classification at 6 Hz ($p < 0.05$), $r = -0.69$ for CSP classification with the filter bank ($p < 0.01$). Accordingly, brain signals of participants with a high level of alpha activity were less modulated by quality changes. One reason for an increased level of alpha activity is decreased attention.

4. Discussion

EEG data can be analyzed from two main points of view, the one that refers to the modulation of brain rhythms and the one that takes into account the time-locked components of brain activity (Lemm *et al* 2011). In motor imagery-based BCIs a well established method for signal processing is the CSP filtering, which finds spatial filters that discriminate the areas of the motor cortex where the modulation of the oscillatory idle state is the highest (Blankertz *et al* 2008). In ERP-based BCIs instead, since the ERPs are time locked to the target event to classify, it is crucial to identify the time intervals in which the discrimination between the targets and non-targets

is the highest. The classification of spatio-temporal features using LDA with a regularization by shrinkage of the covariance matrix was found to be successful (Blankertz *et al* 2011, Bartz and Müller 2013). SSVEPs are evoked potentials time-locked to the attended stimulus, which differ from ERPs by being in steady state and not transient. In most of the BCIs studies based on SSVEPs the EEG signal is processed in the frequency domain because usually the stimuli have different frequencies in order to differentiate several target events. In our case, we adopted one frequency for the stimulation, since our purpose was not to operate a BCI (Cheng *et al* 2002, Müller-Putz *et al* 2005, Allison *et al* 2008) but to modulate the EEG activity according to the quality change of the stimuli. We can still exploit the oscillatory nature of the signal with a narrow band-pass filtering around the frequencies in which the variance of the signal is enhanced compared to the background state. But also the temporal evolution of the evoked potential can be investigated, since the SSVEPs are time locked to the quality changes. The two methods used in this study for offline analysis take into account both phenomena, and lead to results with comparable trends but different accuracies. Classification results show mean offline accuracies up to 0.74 and 0.84 for the maximum distortion level D6, respectively for the CSP method and the spatio-temporal features method. Both methods show the same trend in the classification results, that is, the performances do not exceed chance level significantly for the first three levels of distortion, while from D4 to D6 they increase significantly and linearly with the quality change. These results suggest that the quality changes introduced in the textures by the compression algorithm modulate the ongoing EEG activity eliciting SSVEPs which can be classified over the perception threshold. The accuracies achieved in the proposed design are in the same range of those exploiting the P3 component as neural feature for video quality assessment via EEG. Note that, performances between these two types of studies are very difficult to compare because of the different design of the stimuli used in P3-based and SSVEP-based experiments. Mustafa *et al* (2012) reach a mean single-trial classification accuracy up to 85% in classifying the presence of artifacts in videos versus ground truth, using a wavelet-based classification. The most severe artifacts reach a mean detection of 93%. Lindemann *et al* (2011) also perform single trial classification of the artifact in videos vs ground truth, using principal component analysis (PCA) for dimensionality reduction and support vector machine (SVM) as classifier. They use only the trials correctly detected by the participants via button press, achieving a mean classification accuracy of 76.5% for the most obvious artifacts and 73.5% for the less obvious. Scholler *et al* (2012) report a single-trial classification with AUC values close to 1 for the highest level of distortion in most the subjects. This result is obtained after filtering raw data with a LDA filter, in which the weights are computed based on the signed squared biserial correlation coefficient between the trials with the highest quality change and the trials without quality change. The classification accuracies reported by Scholler and Lindemann refer only to trials correctly identified by the participants at the behavioral level.

Scholler *et al* also report for three participants an average 65% accuracy in classifying the trials in which the quality change was present but not detected by the subjects, advancing the hypothesis of higher sensitivity of the EEG compared to the behavioral response. Since our study capitalizes on SSVEPs elicited in condition of passive viewing, we could not differentiate between trials potentially ‘detected’ by the participants and trials in which participants might have had a lower level of attention. So our accuracies refer to the overall number trials, without pre-screening or rejections. About the number of trials, it has to be pointed out that a SSVEP-based paradigm allows the recording of an amount much higher compared to P3-based paradigms. In fact, in the latter the number of epochs containing the actual target event is just a fraction of the total amount of epochs used in the experiment. In our design, each epoch is a ‘target’, meaning that all the trials contain useful information for the quality assessment. As already mentioned in the introduction, the number of events (as event referring to the occurrence of evoked potentials useful for the evaluation) is increased more than ten times if compared to the P3-based study of Scholler *et al* (2012). That is to say that a SSVEP-based assessment can be more than ten times faster than a P3 one. If we want to quantify the speed of detection in terms of bit rate (Wolpaw *et al* 1998), our system reaches about 27 bits min⁻¹ for classification of D6 with the spatio-temporal features method. This performance is achieved using just 13 min of EEG recording for the training of the classifier.

Another fundamental aspect to consider is the comparison between the neural assessment and the behavioral one. Studies which foresee participants pressing a button at detection of artifacts or distortions allow a more straightforward comparison between the neural and the behavioral response (Porbadnigk *et al* 2013). However, in these cases there is no indication of the actual level of quality perceived by the participants, that is, how well they could detect the distortions and how annoying the detected artifact could be. And this is actually an important detail to take into account in the implementation of video codecs. Therefore, unlike most of previous works, we performed the behavioral assessment in a separate test in which we collected the MOS values that represent the actual subjective rate normally used in image quality assessment. We then linked these results to the accuracies of the classification of the SSVEPs modulated by the quality changes. This represents a key factor in the design of the paradigm, since previous studies usually refer the classification performances to the presence of an artifact or correlate the results with the absolute value of the quality change. Also in Mustafa *et al* (2012) participants performed the assessment via a MOS scale after each video, but they do not report the correlation with the classification performances. As already mentioned in the introduction, Arndt *et al* (2014) correlate the P3 amplitudes with the MOS values of the participants. They found an average significant correlation in three experiments, in which video quality degradation is caused by artificial blockiness. Anyway, this correlation is not reported in the experiment in which the authors investigate a more realistic scenario, using a real existing codec for

introducing video distortions. In our case, we could prove a significant correlation between the neural and the behavioral assessment using the MOS values. This result is valid for all participants for the spatial-temporal features classification, and for all except two for the CSP based classification. The two participants who show no correlation also have very low performances in general in CSP classification, both with maximum accuracy of 0.57 at D6. The spectra of these two participants also do not display evident peaks at 3 and 6 Hz until level D5. This result suggests that for those participants the spectral features might not be very informative and the spatio-temporal feature classification is preferable (classification accuracy at D6 of 0.76 and 0.75, respectively). In both the methods we used, mean classification remains at chance level for distortions D1, D2, D3 and increases linearly and significantly from D4 to D6. In general, the same trend is visible in most of the participants. Clearly, not all participants are expected to be sensitive in the same manner to the flicker of the quality change. For example, considering the classification method based on spatio-temporal features it can be noticed that five participants have a classification accuracy which stays at chance level at D4, a distortion level which is supposed to be above perception threshold. Even though also for these participants accuracy increases significantly with increasing distortion, at D6 it does not go beyond 75%. For these participants the SSVEPs are less pronounced in the time domain, as well as the spectra peaks at the discriminative frequencies. They seem to be 'less sensitive' to the quality changes in general. There might be different reasons for such phenomenon, which lie in the nature of perception itself. One further reason can be found in the results of the analysis of the occipital alpha rhythm during the experiment. A significant negative correlation was found between the classification performances and the mean magnitude of the alpha peak in the occipital cortex, evaluated at electrode Oz. In literature, the relationship between the spontaneous oscillatory activity when the stimulus is presented and the perception of the stimulus itself has been widely investigated. In particular, different studies (Brandt and Jansen 1991, Jansen and Brandt 1991, Ergenoglu *et al* 2004, Hanslmayr *et al* 2005, 2007, Romei *et al* 2008, Busch *et al* 2009, Busch and VanRullen 2010) show that oscillations in the alpha frequency band interfere with the processing of the visual information and modulate the gain of the visual system. A decreased oscillatory activity is thought to reflect a state of enhanced cortical excitability, and increased activity to reflect a state of cortical idling or inhibition in which excitability is reduced. Our results show that the magnitude of the alpha-power during the experiment significantly correlates negatively with the classification accuracies, that directly reflect the trend of the SSVEPs. This could justify the attenuated modulation and poor classification performances of some participants. The high alpha rhythm during the experiment could have kept their occipital cortex in an idle state preventing a strong modulation by the flickering of the quality change. Despite not showing a pronounced neural modulation at high levels of quality changes (for example D4), these participants were nevertheless able to perceive the

corresponding quality changes during the behavioral assessment. Clearly, conclusions on these findings have to be made carefully. First, the procedure of the rating and the EEG experiment are intrinsically different: the behavioral assessment is based on the comparison between the undistorted and the distorted image, which are presented simultaneously in the display, and participants have 10 s to carefully look at them in a free viewing condition. Participants are then asked to use a mouse to select the decided score, moving a scroll bar between one and nine. All these tasks require people to be active and overtly engaged, and maybe more concentrated in recognizing a possible distortion in the texture to evaluate. During the EEG recording, participants were asked to focus straight in the center of the video, limiting any kind of movement. This task is quite tiring and a reason why for some people it could have been much more difficult to concentrate. The state of excessive relax reflected by the high alpha rhythm could prevent them to properly pay attention to the quality changes. In general, it is well known in SSVEP literature that the amplitude of the modulation is substantially increased by attention (Müller *et al* 1998, 2003) and that an attended flickering stimulus elicits a larger steady-state response than the same stimulus when unattended (Ding *et al* 2006, Müller *et al* 2006). This is also the main principle used by SSVEP-based BCIs (Cheng *et al* 2002, Müller-Putz *et al* 2005, Allison *et al* 2008), in which stimuli are presented simultaneously in different locations and frequencies. The user can focus the attention on the target stimulus whose frequency would modulate the steady-state response accordingly. Another reason for the poorer performances of these participants could lie in the choice of the stimulus frequency. For some people not all the harmonics are modulated by attention in the same way. For example, in the study of Pei *et al* (2002) stimuli at 2.4 and 3 Hz were chosen and the harmonic responses at 4.8 and 6 Hz were modulated by attention, while the responses at 9.8 and 12 Hz were not. We had a similar result showing the steady-state response present mainly at 3 and 6 Hz, while the modulation of higher harmonics was negligible. The choice of the stimulation frequency was also addressed in the study of Kelly *et al* (2005). They developed a SSVEP-based BCI in which stimuli were presented at a flickering frequency within and outside the alpha range, respectively. Results show that no advantage is gained by using one or the other solution when working with SSVEP features, but the authors suggest that the choice of the stimulus frequency should be specific to individual subjects. In our study, the frequency was chosen beforehand and kept constant for all the subjects. Some of them might have been less sensitive to that specific frequency, no matters their engagement in the task or the idle state of their alpha rhythm. Some research show a percentage of people being not able to operate an SSVEP-based BCI at all (Allison *et al* 2008, 2010, Volosyak *et al* 2011). This percentage is lower than the estimation of the 'BCI-illiterates' in motor-imaging based BCIs (between 15–30%, Blankertz *et al* (2010), Hammer *et al* (2012), Suk *et al* (2014)), but it is an aspect worthwhile to study in the future.

4.1. Limitation and future developments

The six levels of distortions were associated to the QP which were chosen in a pilot study such that the perceived degradation at the behavioral level was the same for all the textures. In the pilot study we performed a behavioral test like described in the methods section, and finally we chose the QP for each texture such that the average MOS values across subjects was slightly below perception threshold for D2 and slightly above for D3. In the main study we could not reproduce the same results for a new pool of subjects. On average, both D2 and D3 were below the perception threshold both on the behavioral level and on the EEG level, therefore not discriminable by the classifier. Future studies could take into account more levels of the QP, especially increasing the number of the levels around the perception threshold in order to investigate the sensitivity of the EEG in relationship with the behavioral results. In our results we did not find an evidence of higher sensitivity of the EEG compared to the overt response, suggesting that there were no distortions processed unconsciously which would not result at the behavioral level. Previous ERP studies (Porbadnigk *et al* 2011, Scholler *et al* 2012) suggest that some people might show an unconscious neural processing of quality changes. Therefore, it is crucial to investigate more in depth the relationship between EEG response and behavioral one at the perception threshold, in order to prevent the implementation of video coding methods which would introduce distortions potentially perceived by the most sensitive people. As already mentioned, the stimuli presentation in the behavioral test and in the EEG part are intrinsically different. In the EEG part, participants had to attend to a video in which the textures were presented in succession with a quality changes at a frequency of 3 Hz and in the behavioral assessment they had to rate the quality of the same textures displayed together with the reference image (following the video quality assessment standard tests performed according to ITU (2002)). A future study could consider to present the images to be assessed behaviorally in a way that would be more consistent with the stimuli presentation in the EEG part. For example, the reference image and the distorted one can be presented right after each other for several seconds before the rating. The behavioral assessment could be made after each video, helping participants to be more awake and concentrated during the recordings. We showed the correlation between the perceived quality change and the size of the neural effect. In this study image degradation was caused by a specific compression algorithm, which introduces changes in more fundamental image features. The identification of which specific feature is responsible of the generation of the SSVEPs was beyond the scope of the paper. This leads to the conclusion that the applicability of this method to other image compression algorithms has to be considered with caution, since other algorithms may control the image features in a different way.

5. Conclusions

We showed that the quality changes introduced in natural textures by the HM10.0 test model of the H.265/MPEG-HEVC standard could be measured by EEG. For that, the distorted signal modulated the neural signal, eliciting SSVEPs that could be classified with high accuracies over the perception threshold. The proposed experimental design let us collect a number of epochs an order of magnitude higher than previous P3-based designs, in the same period of time. The results of the neural assessment significantly correlated with the MOS values of the behavioral assessment. Taking into account the characteristic of the visual system which sees people having different sensitivity to the stimuli, this design could be further improved choosing subject-specific stimulation frequencies, increasing the number of the assessed quality levels, and monitoring the level of the alpha rhythm during the experiment. In general, one can conclude that assessing video quality via the SSVEP-based paradigm is not only a useful complement to the standard behavioral tests but also a significantly faster alternative to P3-based paradigms.

Acknowledgments

We acknowledge financial support by the BMBF Grant N. 01GQ0850.

References

- Allison B, Luth T, Valbuena D, Teymourian A, Volosyak I and Graser A 2010 Bci demographics: how many (and what kinds of) people can use anssvep bci? *IEEE Trans. Neural Syst. Rehabil. Eng.* **18** 107–16
- Allison B Z, McFarland D J, Schalk G, Zheng S D, Jackson M M and Wolpaw J R 2008 Towards an independent brain–computer interface using steady state visual evoked potentials *Clin. Neurophysiol.* **119** 399–408
- Ang K K, Chin Z Y, Zhang H and Guan C 2008 Filter bank common spatial pattern (fbcspp) in brain–computer interface *IEEE Int. J. Conf. Neural Networks* pp 2390–7
- Antons J-N, Schleicher R, Arndt S, Moller S, Porbadnigk A and Curio G 2012 Analyzing speech quality perception using electroencephalography *J. Sel. Top. Signal Process. IEEE* **6** 721–31
- Arndt S, Antons J-N, Schleicher R, Moller S and Curio G 2014 Using electroencephalography to measure perceived video quality *J. Sel. Top. Signal Process. IEEE* **8** 366–76
- Babiloni C, Vecchio F, Bultrini A, Romani G L and Rossini P M 2006 Pre-and poststimulus alpha rhythms are related to conscious visual perception: a high-resolution EEG study *Cereb Cortex* **16** 1690–700
- Bartz D M and Müller K-R 2013 Generalizing analytic shrinkage for arbitrary covariance structures *Adv. Neural Inf. Proc. Syst.* **26** 1869–77
- Blankertz B, Lemm S, Treder M S, Haufe S and Müller K-R 2011 Single-trial analysis and classification of ERP components—a tutorial *Neuroimage* **56** 814–25
- Blankertz B, Sannelli C, Halder S, Hammer E M, Kübler A, Müller K-R, Curio G and Dickhaus T 2010

- Neurophysiological predictor of SMR-based BCI performance *Neuroimage* **51** 1303–9
- Blankertz B, Tomioka R, Lemm S, Kawanabe M and Müller K-R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Process. Mag.* **25** 41–56
- Bosse S, Acqualagna L, Porbadnigk A, Blankertz B, Curio G, Müller K-R and Wiegand T 2014 Neurally informed assessment of perceived natural texture image quality *Proc. IEEE Int. Conf. Image Process.* pp 1987–91
- Brandt M E and Jansen B H 1991 The relationship between prestimulus alpha amplitude and visual evoked potential amplitude *Int. J. Neurosci.* **61** 261–8
- Busch N A, Dubois J and VanRullen R 2009 The phase of ongoing eeg oscillations predicts visual perception *J. Neurosci.* **29** 7869–76
- Busch N A and VanRullen R 2010 Spontaneous eeg oscillations reveal periodic sampling of visual attention *Proc. Natl Acad. Sci.* **107** 16048–53
- Calhoun G L and McMillan G R 1996 EEG-based control for human-computer interaction *Proc. 3rd Annual Sym. on Human Inter with Complex Sys. HICS96* pp 4–9
- Cheng M, Gao X, Gao S and Xu D 2002 Design and implementation of a brain-computer interface with high transfer rates *IEEE Trans. Biomed. Eng.* **49** 1181–6
- Demiralp T, Bayraktaroglu Z, Lenz D, Junge S, Busch N A, Maess B, Ergen M and Herrmann C S 2007 Gamma amplitudes are coupled to theta phase in human eeg during visual perception *Int. J. Psychophysiol.* **64** 24–30
- Ding J, Sperling G and Srinivasan R 2006 Attentional modulation of ssvep power depends on the network tagged by the flicker frequency *Cereb Cortex* **16** 1016–29
- Donchin E 1979 Event-related brain potentials: a tool in the study of human information processing *Evok Brain Pot Behav* (Berlin: Springer) pp 13–88
- Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G and Müller K-R 2006 Combined optimization of spatial and temporal filters for improving brain-computer interfacing *IEEE Trans. Biomed. Eng.* **53** 2274–81
- Duncan-Johnson C C and Donchin E 1977 On quantifying surprise: the variation of event-related potentials with subjective probability *Psychophysiology* **14** 456–67
- Ergenoglu T, Demiralp T, Bayraktaroglu Z, Ergen M, Beydagi H and Uresin Y 2004 Alpha rhythm of the eeg modulates visual detection performance in humans *Cogn. Brain Res.* **20** 376–83
- Friman O, Volosyak I and Graser A 2007 Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces *IEEE Trans. Biomed. Eng.* **54** 742–50
- Gao X, Xu D, Cheng M and Gao S 2003 A bci-based environmental controller for the motion-disabled *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 137–40
- Hammer E M, Halder S, Blankertz B, Sannelli C, Dickhaus T, Kleih S, Müller K-R and Kübler A 2012 Psychological predictors of SMR-BCI performance *Biol. Psychol.* **89** 80–86
- Hanley J A and McNeil B J 1982 The meaning and use of the area under a receiver operating characteristic (roc) curve *Radiology* **143** 29–36
- Hanslmayr S, Aslan A, Staudigl T, Klimesch W, Herrmann C S and Bäuml K-H 2007 Prestimulus oscillations predict visual perception performance between and within subjects *Neuroimage* **37** 1465–73
- Hanslmayr S, Klimesch W, Sauseng P, Gruber W, Doppelmayr M, Freunberger R and Pecherstorfer T 2005 Visual discrimination performance is related to decreased alpha amplitude but increased phase locking *Neurosci. Lett.* **375** 64–68
- Hauße S, Meinecke F, Görgen K, Dähne S, Haynes J, Blankertz B and Biessmann F 2014 On the interpretation of weight vectors of linear models in multivariate neuroimaging *Neuroimage* **87** 96–110
- Hayashi H, Shirai H, Kameda M, Kunifuji S and Miyahara M 2000 Assessment of extra high quality images using both eeg and assessment words on high order sensations *IEEE Int. Conf. on Systems, Man and Cybernetics* vol **2** 1289–94
- Herrmann C S 2001 Human eeg responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena *Exp. Brain Res.* **137** 346–53
- Huber P J and Ronchetti E M 1975 Robustness of design *Robust Statistics* 2nd edn (Hoboken, NJ: Wiley) pp 239–48
- ITU 2002 Methodology for the subjective assessment of the quality of television pictures *Rec. ITU-R BT 500-11* (www.itu.int/rec/R-REC-BT.500)
- ITU 2008 Subjective video quality assessment methods for multimedia applications *Rec. ITU-T P.910* (www.itu.int/rec/R-REC-BT.500)
- Jansen B H and Brandt M E 1991 The effect of the phase of prestimulus alpha activity on the averaged visual evoked response *Electroencephalogr. Clin. Neurophysiol., Evoked Potentials Sect.* **80** 241–50
- Jayant N, Johnston J and Safranek R 1993 Signal compression based on models of human perception *Proc. IEEE* **81** 1385–422
- JCT-VC 2014 Subversion repository for the hevc test model reference software (<https://hevc.hhi.fraunhofer.de>)
- Johnson R 1993 On the neural generators of the p300 component of the event-related potential *Psychophysiology* **30** 90–97
- Keil A, Gruber T, Müller M M, Moratti S, Stolarova M, Bradley M M and Lang P J 2003 Early modulation of visual perception by emotional arousal: evidence from steady-state visual evoked brain potentials, Cogn, Affective Behav. *Neurosci.* **3** 195–206
- Kelly S P, Lalor E C, Reilly R B and Foxe J J 2005 Visual spatial attention tracking using high-density ssvep data for independent brain-computer communication *IEEE Trans. Neural Syst. Rehabil. Eng.* **13** 172–8
- Koles Z, Lazar M and Zhou S 1990 Spatial patterns underlying population differences in the background EEG *Brain Topogr.* **2** 275–84
- Kroupi E, Hanhart P, Lee J-S, Rerabek M and Ebrahimi T 2014 EEG correlates during video quality perception *European Sign Process Conf.* pp 2135–9
- Kylberg G 2011 The kylberg texture dataset v. 1.0 *External Report* (Uppsala University Sweden: Centre for Image Analysis) **35**
- Lemm S, Blankertz B, Curio G and Müller K-R 2005 Spatio-spectral filters for improving the classification of single trial eeg *IEEE Trans. Biomed. Eng.* **52** 1541–8
- Lemm S, Blankertz B, Dickhaus T and Müller K-R 2011 Introduction to machine learning for brain imaging *Neuroimage* **56** 387–99
- Lindemann L and Magnor M 2011 Assessing the quality of compressed images using EEG *18th IEEE Int. Conf. Image Processing* pp 3109–12
- Lindemann L, Wenger S and Magnor M 2011 Evaluation of video artifact perception using event-related potentials *Proc. ACM SIGGRAPH Symp. Appl. Percept in Graph and Visualiz* pp 53–58
- Little D 2007 The common european framework of reference for languages: perspectives on the making of supranational language education policy *J. Mod. Lang.* **91** 645–55
- Lotte F et al 2007 A review of classification algorithms for eeg-based brain-computer interfaces *J. Neural Eng.* **4** 1–13
- Moldovan A-N, Ghergulescu I, Weibelzahl S and Muntean C 2013 User-centered eeg-based multimedia quality assessment *IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting* pp 1–8
- Morgan S, Hansen J and Hillyard S 1996 Selective attention to stimulus location modulates the steady-state visual evoked potential *Proc. Natl Acad. Sci.* **93** 4770–4

- Müller M, Andersen S, Trujillo N, Valdes-Sosa P, Malinowski P and Hillyard S 2006 Feature-selective attention enhances color signals in early visual areas of the human brain *Proc. Natl Acad. Sci.* **103** 14250–4
- Müller M M and Hübner R 2002 Can the spotlight of attention be shaped like a doughnut? evidence from steady-state visual evoked potentials *Psychol. Sci.* **13** 119–24
- Müller M, Malinowski P, Gruber T and Hillyard S 2003 Sustained division of the attentional spotlight *Nature* **424** 309–12
- Müller M M, Picton T W, Valdes-Sosa P, Riera J, Teder-Sälejärvi W A and Hillyard S A 1998 Effects of spatial selective attention on the steady-state visual evoked potential in the 20–28 Hz range *Cogn. Brain Res.* **6** 249–61
- Müller-Putz G R, Scherer R, Brauneis C and Pfurtscheller G 2005 Steady-state visual evoked potential (ssvep)-based communication: impact of harmonic frequency components *J. Neural Eng.* **2** 123
- Mustafa M, Guthe S and Magnor M 2012 Single-trial EEG classification of artifacts in videos *ACM Trans. Appl. Percept.* **9** 12:1–12:15
- Norcia A, Ales J, Cooper E and Wiegand T 2014 Measuring perceptual differences between compressed and uncompressed video sequences using the swept-parameter visual evoked potential *J. Vision* **14** 649–649
- Ojala T, Maenpää T, Pietikainen M, Viertola J, Kyllönen J and Huovinen S 2002 Outex-new framework for empirical evaluation of texture analysis algorithms *Conf. on Pattern Recognition, 2002. Proc. 16th Int. vol 1* (Piscataway, NJ: IEEE) 701–6
- Parini S, Maggi L, Turconi A C and Andreoni G 2009 A robust and self-paced bci system based on a four class ssvep paradigm: algorithms and protocols for a high-transfer-rate direct brain communication *Comput. Intell. Neurosci.* **2009** 864564
- Pastor M A, Artieda J, Arbizu J, Valencia M and Masdeu J C 2003 Human cerebral activation during steady-state visual-evoked responses *J. Neurosci.* **23** 11621–7
- Pei F, Pettet M W and Norcia A M 2002 Neural correlates of object-based attention *J. Vision* **2** 1
- Perez J and Delecchelle E 2013 On the measurement of image quality perception using frontal eeg analysis *Int. Conf. on Smart Communications in Network Technologies (SaCoNeT)* **01** 1–5
- Pfurtscheller G and Lopes da Silva F H 1999 Event-related eeg/meg synchronization and desynchronization: basic principles *Clin. Neurophysiol.* **110** 1842–57
- Picton T W 1992 The p300 wave of the human event-related potential *J. Clin. Neurophysiol.* **9** 456–79
- Picton T et al 2000 Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria *Psychophysiology* **37** 127–52
- Porbadnigk A K, Antons J-N, Blankertz B, Treder M S, Schleicher R, Möller S and Curio G 2010 Using ERPs for assessing the (sub) conscious perception of noise *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **vol 2010** pp 2690–3
- Porbadnigk A K, Görnitz N, Kloft M and Müller K-R 2013 Decoding brain states during auditory perception by supervising unsupervised learning *J. Comput. Sci. Eng.* **7** 112–21
- Porbadnigk A K, Scholler S, Blankertz B, Ritz A, Born M, Scholl R, Müller K-R, Curio G and Treder M S 2011 Revealing the neural response to imperceptible peripheral flicker with machine learning *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2011** pp 3692–5
- Porbadnigk A K, Treder M S, Blankertz B, Antons J N, Schleicher R, Möller S, Curio G and Müller K-R 2013 Single-trial analysis of the neural correlates of speech quality perception *J. Neural Eng.* **10** 056003
- Rager G and Singer W 1998 The response of cat visual cortex to flicker stimuli of variable frequency *Eur. J. Neurosci.* **10** 1856–77
- Ramoser H, Müller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial eeg during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–6
- Regan D 1989 *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (New York: Elsevier)
- Romei V, Rihs T, Brodbeck V and Thut G 2008 Resting electroencephalogram alpha-power over posterior sites indexes baseline visual cortex excitability *Neuroreport* **19** 203–8
- Samek W, Vidaurre C, Müller K-R and Kawanabe M 2012 Stationary common spatial patterns for brain–computer interfacing *J. Neural Eng.* **9** 026013
- Sannelli C, Vidaurre C, Müller K-R and Blankertz B 2011 Common spatial pattern patches—an optimized filter ensemble for adaptive brain–computer interfaces *J. Neural Eng.* **8** 025012
- Scholler S, Bosse S, Treder M S, Blankertz B, Curio G, Müller K-R and Wiegand T 2012 Towards a direct measure of video quality perception using EEG *IEEE Trans. Image Process.* **21** 2619–29
- Sellers E W, Krusienski D J, McFarland D J, Vaughan T M and Wolpaw J R 2006 A p300 event-related potential brain–computer interface (bci): the effects of matrix size and inter stimulus interval on performance *Biol. Psychol.* **73** 242–52
- Seshadrinathan K and Bovic A C 2010 Motion tuned spatio-temporal quality assessment of natural videos *IEEE Trans. Image Process.* **19** 335–50
- Smith M E, Halgren E, Sokolik M, Baudena P, Musolino A, Liegeois-Chauvel C and Chauvel P 1990 The intracranial topography of the p3 event-related potential elicited during auditory oddball *Electroencephalogr. Clin. Neurophysiol.* **76** 235–48
- Suk H-I, Fazli S, Mehnert J, Müller K-R and Lee S-W 2014 Predicting bci subject performance using probabilistic spatio-temporal filters *PLoS One* **9** e87056
- Sullivan G J, Ohm J, Han W-J and Wiegand T 2012 Overview of the high efficiency video coding (hevc) standard *IEEE Trans. Circuits Syst. Video Technol.* **22** 1649–68
- Tomioka R and Müller K-R 2010 A regularized discriminative framework for eeg analysis with application to brain–computer interface *Neuroimage* **49** 415–32
- Vialatte F-B, Maurice M, Dauwels J and Cichocki A 2010 Steady-state visually evoked potentials: focus on essential paradigms and future perspectives *Prog. Neurobiol.* **90** 418–38
- Volosyak I, Valbuena D, Luth T, Malechka T and Graser A 2011 Bci demographics ii: how many (and what kinds of) people can use a high-frequency ssvep bci? *IEEE Trans. Neural Syst. Rehabil. Eng.* **19** 232–9
- Wolpaw J R, Ramoser H, McFarland D J and Pfurtscheller G 1998 EEG-based communication: improved accuracy by response verification *IEEE Trans. Rehabil. Eng.* **6** 326–33